

Análise da Topologia dos Sistemas Autônomos: Evolução e Predição

Rodrigo Augusto Alves¹, Cristina Duarte Murta²

¹Departamento de Informática, UFPR

²Departamento de Computação, CEFET-MG

Resumo. *A Internet tem evoluído rapidamente, com alterações significativas em suas medidas e características que afetam sua topologia. Este artigo apresenta uma caracterização da evolução da topologia da Internet, no nível dos sistemas autônomos, no período de 2004 a 2007. A evolução de várias métricas é modelada e extrapolações dos modelos são feitas para o período de 2008 a 2010. Os resultados confirmam tendências apontadas na literatura, indicando que a rede está mais compacta e mais conectada, e revelam que a topologia experimenta uma nova fase em sua evolução, com crescimento menos acelerado.*

Abstract. *The Internet has evolved rapidly, revealing important changes in its measurements and features that affects its topology. In this paper we present a characterization of the evolution of the autonomous system network topology in the period from 2004 to 2007. We track the evolution of some topology metrics, proposing models and performing extrapolations for the period from 2008 to 2010. The results confirm trends that have been previously pointed out in the literature. The network gets more compact and more connected. We also present evidence that the Internet topology experiences a new phase in its evolution, showing a less accelerated growth rate.*

1. Introdução

A evolução constante dos sistemas requer atividade contínua de caracterização para revelar seus padrões e identificar características passageiras e perenes ao longo do seu ciclo de vida. A Internet é um sistema que tem evoluído rapidamente, com alterações significativas em várias de suas medidas e características, incluindo sua topologia. O conhecimento obtido na análise da evolução da topologia da Internet é útil em várias áreas, por exemplo, no projetos de novas redes [GENI 2007, FIND 2007], no planejamento de expansão da rede atual, na avaliação de propostas de novos protocolos e aplicações, e no estudo das inter-relações entre a evolução da Internet e aspectos econômicos e sociais do mundo atual.

A topologia da Internet pode ser descrita em dois níveis, no nível dos roteadores e no nível dos Sistemas Autônomos (SAs). Este artigo apresenta uma caracterização da evolução da topologia da Internet, no nível dos sistemas autônomos, no período de 2004 a 2007. Esta caracterização é feita pela análise da evolução de várias métricas ao longo do período. As métricas escolhidas descrevem aspectos essenciais da rede como sua composição em termos de número de nodos e arestas, a caracterização das bordas e da distância entre os nodos. Para cada métrica analisada é proposto um modelo matemático baseado em análise numérica para descrever o comportamento da métrica no período. Os

modelos propostos permitem a extrapolação das curvas e a realização de projeções sobre a evolução dessas métricas no período de 2008 a 2010. Os resultados são comparados com a literatura.

Os resultados indicam que a rede tem crescido de forma menos acelerada do que o observado em períodos anteriores, o que pode ser uma evidência de que ela se encontra em uma nova fase de sua evolução. Um aspecto essencial é que o crescimento do número de conexões é muito maior do que o de nodos, o que torna a rede mais compacta e densa a cada dia. As principais contribuições deste artigo são: (i) uma caracterização da evolução da rede de sistemas autônomos da Internet em período recente; (ii) a proposta de modelos numéricos, para várias métricas, que descrevem a evolução da topologia neste período; (iii) a projeção dos dados para inferir valores das métricas no período de 2008 a 2010.

Este artigo está organizado em seis seções. A seção seguinte apresenta e discute os trabalhos relacionados. A Seção 3 apresenta a base de dados da topologia real utilizada nesse estudo. A Seção 4 apresenta resultados da análise da evolução da topologia no período entre 2004 e 2007. A Seção 5 apresenta modelos para a evolução e resultados de previsão feitos para o período de 2008 a 2010. A última seção apresenta as conclusões do trabalho.

2. Trabalhos Relacionados

Vários trabalhos contribuíram para revelar medidas e características da topologia dos sistemas autônomos da Internet [Park et al. 2004, Chang et al. 2004, Alderson et al. 2005, Mahadevan et al. 2006b, Pastor-Satorras and Vespignani 2004]. Um exemplo da importância da caracterização é a descoberta da grande variabilidade na distribuição dos graus dos vértices da rede [Faloutsos et al. 1999], resultado que teve forte influência em várias pesquisas [Tangmunarunkit et al. 2002, Palmer and Steffan 2000, Jaiswal et al. 2004, Mahadevan et al. 2006a, Bu and Towsley 2002, Chen et al. 2002].

Um grande conjunto de métricas pode ser utilizado para caracterizar a topologia dos sistemas autônomos da Internet, assim como a topologia de várias redes complexas [Pastor-Satorras and Vespignani 2004, Dorogovtsev and Mendes 2003, Alderson et al. 2005, Mahadevan et al. 2006b]. Não há consenso na comunidade sobre quais são as métricas mais importantes e representativas da topologia. As métricas mais analisadas são a distribuição dos graus dos vértices do grafo que representa a topologia, as distâncias entre os vértices e métricas derivadas, e métricas de conectividade.

Coletar dados da topologia da Internet é uma tarefa difícil. A rede não provê informação sobre sua própria topologia e, em geral, essa informação é capturada com o uso de ferramentas e protocolos não destinados a esse fim [Chang and Willinger 2006, Cohen and Raz 2006]. Em consequência, parte da topologia pode não ser capturada, o que sugere que os grafos gerados podem ser incompletos [Zhang et al. 2005, Oliveira et al. 2007].

Várias pesquisas têm sido realizadas com o objetivo de reunir a maior quantidade de dados pertinentes à topologia da Internet no nível dos sistemas autônomos [Chang et al. 2004, Chang and Willinger 2006, He et al. 2007, Mahadevan et al. 2006b, Zhang et al. 2005], utilizando diferentes métodos, tecnologias e ferramentas. A utilização de vários métodos de captura de informações sobre o roteamento é uma das abordagens mais utilizadas atualmente.

O trabalho mais próximo ao presente artigo é apresentado em [Magoni and Pansiot 2001]. Os autores daquele trabalho apresentam uma caracterização detalhada da topologia da Internet, a partir de bases de dados com informações de roteamento BGP, coletadas no período entre novembro de 1997 e maio de 2000. O estudo avalia a topologia a partir de várias métricas e argumenta sobre a evolução da topologia no futuro, propondo leis empíricas sobre métricas que se mostraram invariantes ou de comportamento previsível durante o período analisado, algumas ainda válidas para a rede atual.

O presente trabalho apresenta um estudo similar àquele, porém baseando-se em topologias coletadas no período de 2004 a 2007. Observamos que a Internet está em evolução constante, e a caracterização da topologia deve ser realizada frequentemente. Além disso, a coleta de dados realizada atualmente na rede é mais completa e precisa, o que contribui para revelar informações mais exatas da topologia. Nossos resultados são comparados aos apresentados naquele artigo.

3. Base de Dados da Topologia dos Sistemas Autônomos

A base de dados escolhida para o presente estudo foi elaborada pelo *Internet Research Laboratory* da *University of California, Los Angeles* [IRL 2006], e é denominada IRL nesse trabalho. Sua descrição é apresentada em [Zhang et al. 2005]. A escolha da base de dados da topologia foi baseada nos critérios de completude, atualização e facilidade de acesso. A base IRL é atualizada diariamente, e contém séries históricas acumuladas desde janeiro de 2004 até o presente. O acesso é fácil e essa é a base mais completa disponível [Zhang et al. 2005], contendo cerca de 40% mais arestas e 5% mais vértices em comparação com a base de dados capturada pelo projeto *BGP Route Views* [RouteViews 2007].

Para o estudo proposto nesse artigo, utilizamos dados da base IRL coletados no período de 2004 a 2007. A data de início se deve ao fato de que a base foi coletada a partir deste período, e a data de fim se refere a última disponível no momento da análise dos dados. Observamos que seria interessante analisar períodos anteriores a 2004. Porém, isso somente poderia ser feito utilizando outras bases de dados. No entanto, a diferença entre os métodos de coleta de cada base produz grafos com padrões distintos, o que poderia interferir na análise. Por isso, optamos por analisar dados de uma única base.

4. Evolução da Topologia no Período de 2004 a 2007

Essa seção apresenta uma caracterização da evolução da topologia dos sistemas autônomos no período entre 2004 e 2007. Para cada ano foram coletadas duas topologias, uma em 30 de janeiro e outra em 30 de julho. Portanto, esse estudo analisa oito topologias coletadas de forma regular no período de quatro anos, caracterizando a evolução semestral da Internet quanto à topologia dos sistemas autônomos.

A topologia da rede é representada na forma de um grafo $G = (V, E)$ no qual V representa o conjunto de vértices e E é o conjunto de arestas do grafo. Nesse grafo, cada vértice representa um sistema autônomo (SA) e as arestas representam conexões lógicas entre SAs. As métricas utilizadas para análise da evolução são a quantidade total de nodos e arestas, o número de componentes do grafo, a presença e caracterização de árvores nas

bordas da rede, métricas de distância e derivadas, e métricas relacionadas aos graus dos vértices. Estas métricas caracterizam de forma geral a topologia da rede.

4.1. Composição do Grafo e Graus dos Vértices

A primeira análise contempla a evolução do tamanho da rede, representada pelo número de vértices e de arestas, e a caracterização estatística de uma métrica essencial nesse grafo, o grau dos vértices. A Tabela 1 apresenta os resultados dessa análise para todo o período. As estatísticas apresentadas são a mediana e a média dos graus dos vértices, o maior grau observado em cada topologia e o coeficiente de variação (COV) dos graus dos vértices. O COV é a razão entre o desvio padrão e a média, e é uma medida da variabilidade de um conjunto de dados.

Observa-se que o número de sistemas autônomos quase dobrou no período, enquanto o número de conexões lógicas entre esses sistemas foi multiplicado por quatro. Este resultado indica que a conectividade entre os nodos da rede aumentou em proporções muito maiores do que o tamanho da rede. Essa observação é confirmada pela análise das estatísticas dos graus dos vértices. O grau médio dobrou no período, a mediana passou de 2 para 3, e o maior grau aumentou consistentemente a cada observação. O coeficiente de variação manteve-se aproximadamente constante, em torno de 6, com exceção da primeira observação. Isto indica que a variabilidade dos graus dos vértices não aumentou no período analisado.

Data	Vértices	Arestas	Mediana	Média	Maior	COV
01/2004	16573	39143	2	4,72	2388	4,72
07/2004	18783	62305	2	6,63	2917	6,37
01/2005	20830	76976	2	6,36	3205	6,22
07/2005	23062	92910	2	8,05	3505	6,07
01/2006	24971	106141	3	8,50	3710	6,02
07/2006	27071	121045	3	8,94	3908	5,96
01/2007	29167	136888	3	9,38	4122	5,90
07/2007	31328	156871	3	10,01	4989	5,99

Tabela 1. Composição dos grafos e estatísticas dos graus dos vértices

A Figura 1 apresenta o número de vértices e de arestas no período de observação. O gráfico à esquerda apresenta o crescimento de cada métrica ao longo do período e o gráfico à direita correlaciona o crescimento dos vértices e das arestas em eixos logarítmicos. Observa-se nesse gráfico um padrão bastante previsível para o crescimento da rede.

A evolução da mediana e da média dos graus dos vértices é apresentada no gráfico da Figura 2. No período de observação, a mediana apresentou uma pequena alteração, de 2 para 3, enquanto o grau médio aumentou de aproximadamente 5 para 10. Esse aumento pode ser explicado pelo crescimento dos graus dos nodos mais conectados. O aumento da mediana indica pequeno aumento no grau dos nodos menos conectados. O aumento do grau médio é observado como uma característica da evolução de várias redes complexas [Leskovec et al. 2007].

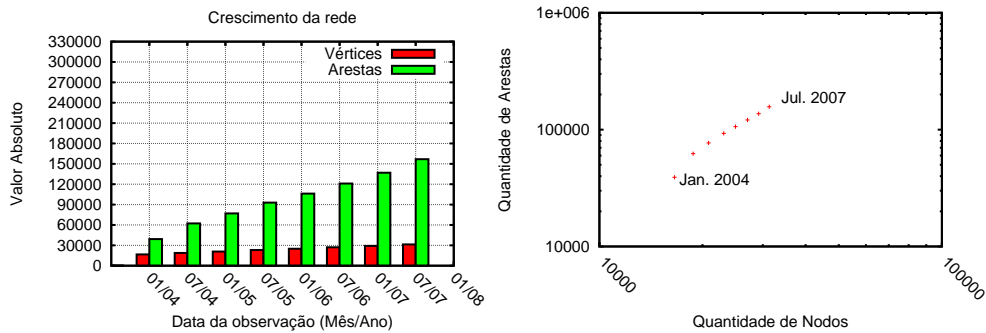


Figura 1. Crescimento da rede no período de 2004 a 2007 (esq.) e correlação entre o número de vértices e arestas em escala logarítmica (dir.)

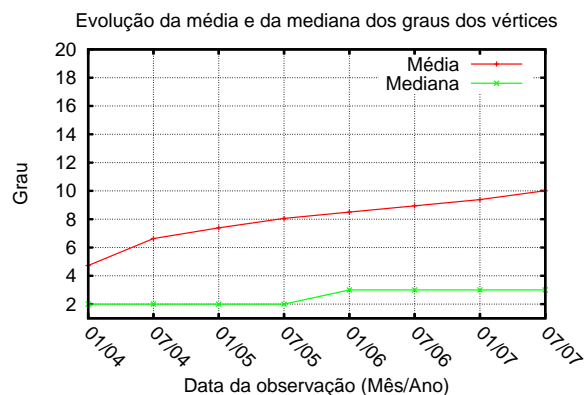


Figura 2. Média e mediana dos graus dos vértices no período de 2004 a 2007

4.2. Distância entre Pares de Vértices

A métrica distância é definida pelo comprimento, em número de arestas, do menor caminho entre pares de vértices. Muitos aspectos do desempenho de protocolos de roteamento e de aplicações dependem da distribuição das distâncias entre pares de vértices na topologia da rede. Em particular, distâncias pequenas e pouco variáveis são desejáveis em algoritmos de roteamento. Porém, distâncias pequenas tornam a rede mais vulnerável a ataques baseados em propagação de agentes. A distância caracteriza também o efeito denominado *small world*, que é o fato de que os pares de vértices estão conectados por caminhos bem curtos.

A Tabela 2 apresenta estatísticas da distância entre pares de vértices. Observa-se que os pares de vértices estão a distâncias muito pequenas e pouco variáveis. A maior distância oscilou entre 9 e 10 arestas no período. A distância média está entre 3 e 4 arestas, o que evidencia os caminhos curtos entre pares de vértices. A distância média e o coeficiente de variação reduziram de forma consistente no período de observação, o que indica aumento global na conectividade da rede, observado de forma genérica, e não apenas para subconjuntos de vértices. Esses resultados são compatíveis com os apresentados em [Magoni and Pansiot 2001].

As frequências absoluta e relativa das distâncias entre pares de vértices são apresentadas na Figura 3. Observa-se um crescimento bem regular das distâncias entre pares de vértices no período. Os gráficos mostram que em janeiro de 2004 havia maior

Data	Média	Maior	COV
01/2004	3,708	10	0,241
07/2004	3,561	9	0,230
01/2005	3,531	9	0,230
07/2005	3,492	9	0,224
01/2006	3,480	9	0,221
07/2006	3,471	9	0,218
01/2007	3,461	10	0,215
07/2007	3,431	9	0,215

Tabela 2. Evolução das estatísticas das distâncias no período de 2004 a 2007

proporção de distâncias 4 em relação às demais distâncias, fato que se reverteu para a distância 3 no período analisado. Observa-se um aumento consistente na fração de distâncias 3 ao longo do período, e o padrão de maior frequência para a distância 3, seguida da distância 4.

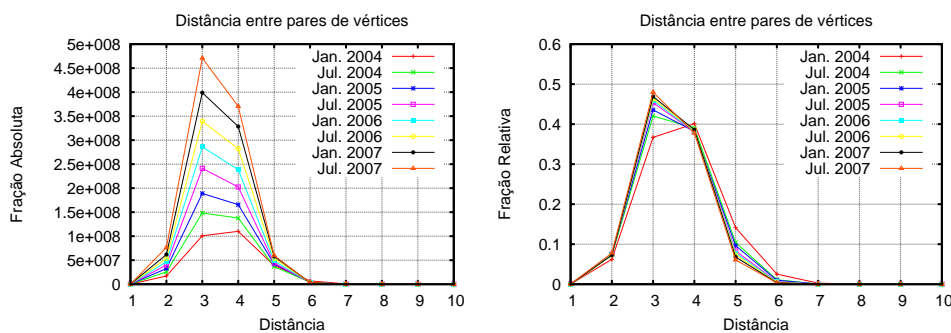


Figura 3. Frequências absoluta (esq.) e relativa (dir.) da distância entre pares de vértices no período de 2004 a 2007

O comportamento das distâncias mais frequentes é analisado na Figura 4. O gráfico à esquerda ilustra a evolução das distâncias 3 e 4 no período observado. A porcentagem de pares de vértices que estão à distância 3 entre si aumentou de cerca de 37% para 48%, enquanto a fração de distâncias 4 caiu de 40% para 38% entre 2004 e 2007. O gráfico à direita da mesma figura apresenta a evolução das distâncias 2, 5 e 6 no período. Enquanto a frequência da distância 2 permaneceu praticamente estável, as ocorrências de distâncias 5 e 6 sofreram queda perceptível.

A análise conjunta da evolução das distâncias indica que a rede, mesmo tendo crescido significativamente durante o período, está se compactando em torno da distância 3. Houve diminuição gradativa da distância média entre 2004 e 2007, de 3,708 para 3,431. Essa compactação da rede, a despeito de seu crescimento em número de vértices, foi observada também em estudos que analisaram topologias nos períodos entre 1997 e 2000 [Magoni and Pansiot 2001, Leskovec et al. 2007] e entre 1999 e 2002 [Bu and Towsley 2002].

Estes resultados indicam que a distância é uma métrica característica da rede, pois apresenta um comportamento bastante regular e invariante. Não obstante a expansão da rede em termos do número de vértices, observa-se um crescimento ainda maior do

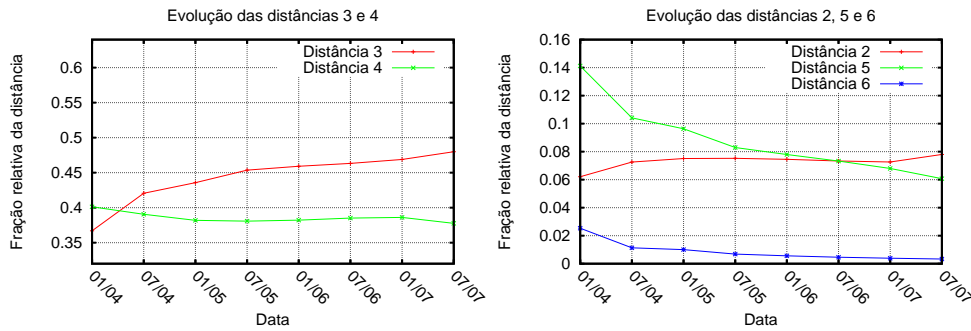


Figura 4. Evolução das frequências das distâncias de 3 e 4 (esq.) e das distâncias 1, 2, 5 e 6 (dir.)

número de conexões entre os vértices, o que explica a diminuição da distância média, que tem decrescido de forma lenta e consistente, tornando a rede mais conexa e compacta.

4.3. Caracterização da Excentricidade

A partir da medida de distância d , definimos a excentricidade. A excentricidade de um vértice v é a distância para o vértice mais longe de v : $e(v) = \max\{d(u, v) | u, v \in V\}$. A excentricidade é uma medida do desvio ou afastamento do centro e indica quão centrais ou periféricos são os vértices. A Tabela 3 apresenta estatísticas relativas à excentricidade nas topologias coletadas ao longo do período. Os vértices mais longínquos estão, em média, à distância 6, e no máximo à distância 9 ou 10.

Data	Raio	Moda	Média	Diâmetro	Centro	Periferia
01/2004	5	7	7,135	10	1	13
07/2004	5	6	6,346	9	173	10
01/2005	5	6	6,360	9	128	15
07/2005	5	6	6,263	9	542	8
01/2006	5	6	6,199	9	1059	2
07/2006	5	6	6,214	9	895	3
01/2007	5	7	6,981	10	6	3
07/2007	5	7	6,760	9	16	67

Tabela 3. Evolução da excentricidade no período de 2004 a 2007

Métricas de centralidade podem ser definidas a partir da excentricidade dos vértices de um grafo. O raio $R(G)$ de um grafo G corresponde à excentricidade mínima de seus vértices: $R(G) = \min\{e(v) | v \in V\}$. O diâmetro $D(G)$ de um grafo G corresponde à excentricidade máxima de seus vértices: $D(G) = \max\{e(v) | v \in V\}$. Além disso, um nodo é centro se $e(v) = R(G)$. O centro de G é o conjunto de vértices de menor excentricidade. Um nodo é periférico se $e(v) = D(G)$. De forma análoga ao centro de G , a periferia de G é o conjunto de vértices do grafo com a maior excentricidade. O resultado para essas métricas é apresentado na Tabela 3.

Para os vértices centrais do grafo, os vértices mais longínquos estão à distância 5, que é o raio do grafo. O diâmetro varia entre 9 e 10, e indica a maior distância entre vértices periféricos. A Tabela 3 apresenta também a quantidade de vértices no centro

e na periferia de cada grafo. Observamos que há maior número de nodos no centro do que na periferia. Ainda assim, o centro e a periferia são compostos por uma pequena fração dos vértices do grafo. Grande parte dos nodos, mais de 90%, está entre o centro e a periferia. Estes resultados indicam que as métricas raio e diâmetro são também invariantes e características da rede.

4.4. Caracterização de Subgrafos Acíclicos

Para analisar a conectividade e as bordas da rede, medimos o número e a altura de subgrafos acíclicos na topologia. Esses subgrafos podem ser entendidos como árvores “penduradas” nas bordas da topologia. A medição foi realizada da seguinte forma: primeiramente, foram retirados todos os vértices de grau 1 e suas respectivas arestas. Em seguida, o mesmo procedimento foi repetido até que mais nenhum vértice de grau 1 fosse encontrado. Para todas as topologias analisadas no período do estudo, de janeiro de 2004 a julho de 2007, apenas duas iterações foram necessárias para eliminar todos os subgrafos acíclicos da rede, mostrando assim que as árvores presentes têm altura máxima igual a 2.

A Figura 5 apresenta a evolução da porcentagem de nodos em árvores na rede, e o tamanho do subgrafo denominado *2-core*, que é o subgrafo que contém apenas nodos de grau 2 ou superior, obtido segundo o procedimento descrito acima. Observamos que a porcentagem de nodos em árvores diminuiu sensivelmente ao longo do período analisado, indicando que o aumento da conectividade ocorre também nas bordas da rede, e não apenas no centro.

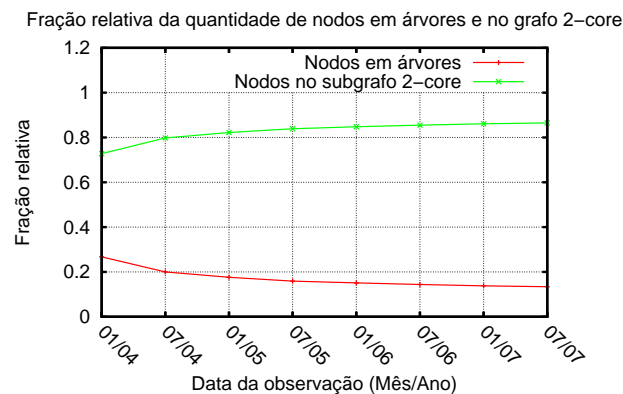


Figura 5. Evolução da fração de nodos em árvores e do tamanho do subgrafo 2-core

5. Previsão da Evolução da Rede

Nesta seção propomos modelos, baseados nos dados históricos, para as métricas descritas na seção anterior. Estes modelos nos permitem fazer extrapolações e inferir medidas futuras da rede. Os modelos foram obtidos utilizando o *Curve Fitting Toolbox* do MATLAB. A ferramenta utiliza várias técnicas de ajuste de curvas e extrapolação. A qualidade dos modelos gerados foi medida pelo intervalo de confiança da medição. Todos os resultados apresentam intervalos de confiança não inferiores a 99%. Cada métrica foi modelada por duas equações que delimitam os valores com base no intervalo de confiança. As curvas plotadas nos gráficos dessa seção foram obtidas tomando-se a média aritmética dos parâmetros das equações que delimitam o conjunto de confiança.

5.1. Previsão do Crescimento da Rede

Para modelar a quantidade de vértices e arestas foram gerados dois polinômios de primeiro grau, um para cada métrica. O objetivo é estimar o crescimento da rede, extrapolando o modelo para o período de 2008 a 2010. A equação obtida para modelar a métrica número de vértices é $14.553 + 2.093x$, obtida a partir das equações $(14.340 + 2.051x; 14.760 + 2.135x)$. Para a medida da quantidade de arestas do grafo, a equação obtida é $27.117 + 15.981x$ com conjunto de confiança igual a $(19.130 + 14.400x; 35.510 + 17.560x)$. A Figura 6 apresenta o modelo produzido, bem como os dados históricos para comparação.

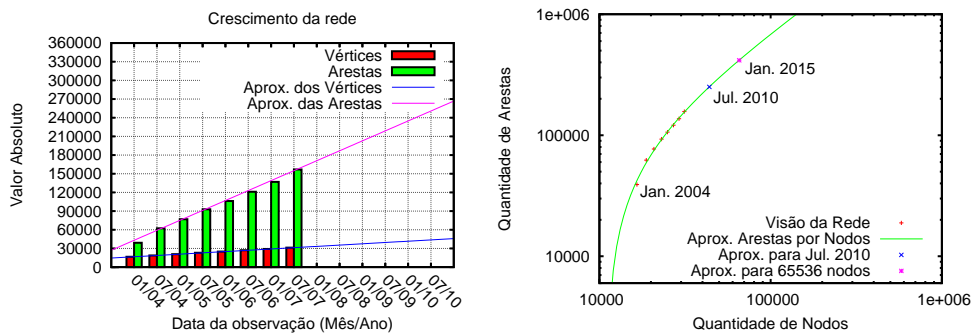


Figura 6. Modelos para o crescimento da rede em quantidade de vértices e arestas no período de 2004 a 2010

Estimativas do crescimento semestral do número de vértices e arestas até o ano de 2010, obtidas a partir do modelo, são apresentadas nas três primeiras colunas da Tabela 4. Quanto ao crescimento da rede, os resultados indicam um padrão bastante diferente do observado em [Magoni and Pansiot 2001]. Leis empíricas enunciadas naquele trabalho indicam crescimento anual de 45% para o número de SAs, e 53% para o número de conexões entre SAs. O padrão atual revela uma taxa de crescimento decrescente, conforme mostra a Figura 7. Cada ponto representa o crescimento da rede em relação ao ponto (semestre) anterior. A rede continua crescendo mas a taxa cada vez menores, o que evidencia uma nova fase em seu ciclo de vida.

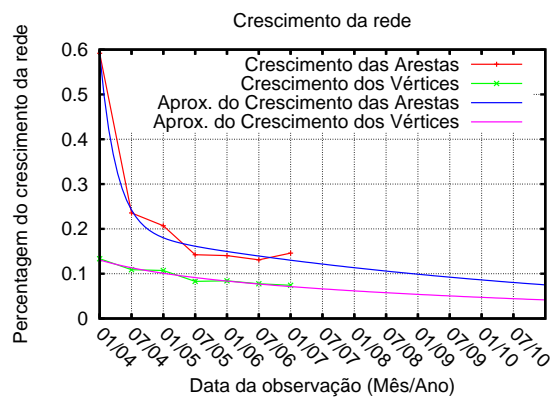


Figura 7. Padrão de crescimento da rede no período de 2004 a 2010

Uma preocupação em relação à expansão da rede é relativa ao esgotamento do espaço de endereçamento dos SAs. Cada SA recebe um número único, e o número

máximo de endereços que o espaço de endereçamento do protocolo IPv4 permite é $2^{16} = 65.536$. Segundo o modelo proposto, o número de SAs chegará a esse limite entre janeiro e julho de 2015. A previsão feita por [Magoni and Pansiot 2001] indicava que o esgotamento dos endereços se daria em janeiro de 2006, o que não ocorreu.

5.2. Previsão da Evolução dos Graus dos Vértices

Data	Vértices	Arestas	Mediana	Média	Maior
01/2008	33392	170951	3	10,98	5015
07/2008	35486	186933	4	11,68	5331
01/2009	37579	202914	4	12,38	5647
07/2009	39672	218896	4	13,08	5963
01/2010	41766	234877	4	13,79	6279
07/2010	43859	250859	4	14,49	6595

Tabela 4. Previsão da composição dos grafos e de medidas dos graus dos vértices

A Tabela 4 apresenta os números inferidos pelos modelos de crescimento da rede no período de 2008 a 2010. As três últimas colunas apresentam projeções relacionadas aos graus dos vértices: mediana, média e maior grau. A média e mediana dos graus dos vértices foram modeladas por equações polinomiais com intervalo de confiança de 99%. A equação polinomial obtida para a média foi $4,962 + 0,665x$, com conjunto de confiança igual a $(3,526 + 0,381x; 6,398 + 0,949x)$. A aproximação da mediana, obtida pela ferramenta utilizada foi igual a $1,643 + 0,190x$ com conjunto de confiança igual a $(0,829 + 0,0293x; 2,457 + 0,352x)$. Para essa métrica, o valor obtido pela equação foi arredondado para o inteiro mais próximo para tornar o modelo mais realístico (nesse conjunto, a mediana é um número inteiro). Os modelos para a média e a mediana dos graus dos vértices são plotados com os dados da topologia real na Figura 8. Caso a evolução se mantenha nos padrões atuais, em 2010 a mediana dos graus dos vértices será 4 e a média estará entre 14 e 15.

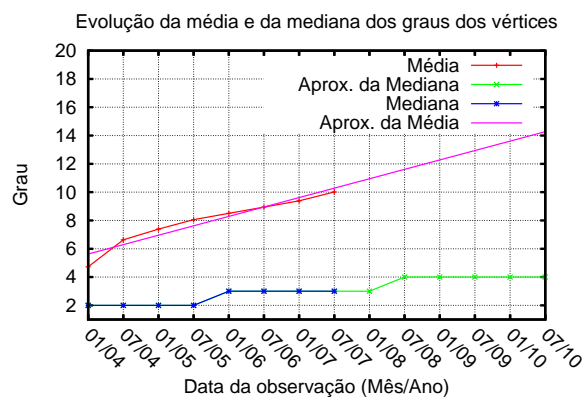


Figura 8. Evolução da média e da mediana dos graus dos vértices no período de 2004 a 2010

Para modelar o maior grau observado, apresentado na última coluna da Tabela 4, foi obtido um polinômio de primeiro grau igual a $4,665 + 0,702x$, representando o con-

junto de confiança $(3,241 + 0,42x; 6,089 + 0,984x)$. Os resultados indicam que, persistindo o padrão de crescimento atual, o grau médio aumentará significativamente, indicando que a rede ficará mais densa, com vértices com graus cada vez maiores.

5.3. Evolução da Distância entre Pares de Vértices

Modelos para a distância média, a maior distância e o coeficiente de variação das distâncias foram também construídos a partir dos dados coletados para essas medidas. O modelo para a distância média é dado pela equação $3,658 - 0,0313x$, um polinômio de primeiro grau, com 99% de confiança e equações limites $(3,528 - 0,0569x; 3,787 - 0,00568x)$. Os valores calculados a partir dos modelos para essas medidas são apresentados na Tabela 5.

Data	Média	Maior	COV
01/2008	3,375	9	0,213
07/2008	3,344	9	0,211
01/2009	3,313	9	0,210
07/2009	3,281	9	0,209
01/2010	3,250	9	0,208
07/2010	3,219	9	0,207

Tabela 5. Estimativas para medidas das distâncias

A distância média entre pares de vértices decresceu no período de janeiro de 2004 a julho de 2007, de 3,708 para 3,431. A extrapolação feita pelo modelo proposto indica que a distância média será próxima de 3,219 em julho de 2010. O coeficiente de variação das distâncias diminuiu de 0,241 para 0,215 entre 2004 e 2007. A extrapolação feita pelo modelo proposto segue a tendência de queda, projetando um valor de 0,207 em julho de 2010.

Para modelar a maior distância, foi escolhida a mediana das maiores distâncias observadas no período de janeiro de 2004 a julho de 2007, cujo valor é igual a 9. Há grande estabilidade nessa medida, e seguindo a tendência observada nos últimos anos, esse valor tende a permanecer estável. A Figura 9 apresenta os modelos obtidos para a métrica de distância.

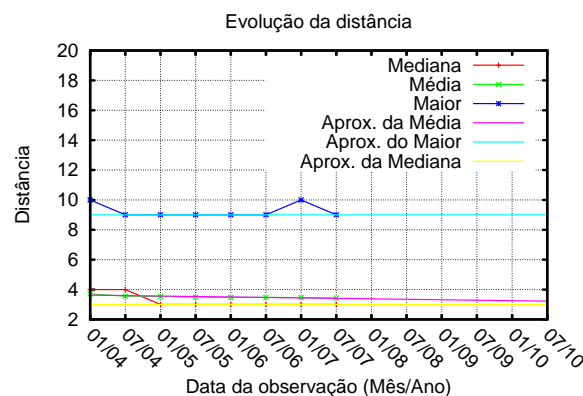


Figura 9. Evolução de medidas das distâncias entre pares de vértices

Os gráficos da Figura 10 apresentam os modelos para a frequência de distâncias 2 a 6, criados a partir dos dados reais analisados (ver Figura 4). As frequências das distâncias 2 e 4 são modeladas por polinômios de primeiro grau com 99% de confiança. Os modelos para a frequência das distâncias 3, 5 e 6 são equações logarítmicas.

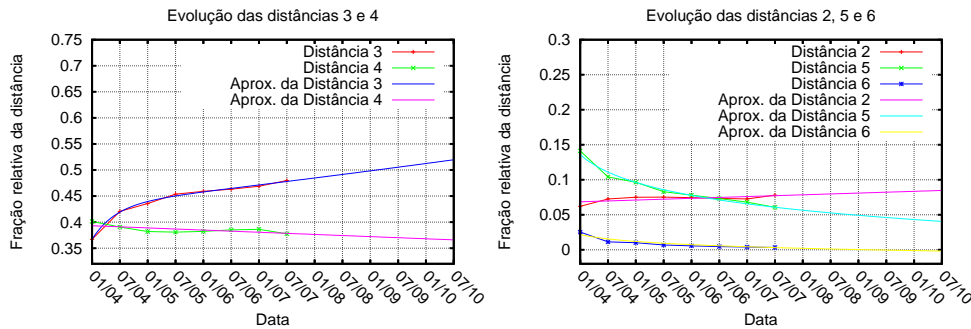


Figura 10. Evolução das distâncias de 3 e 4 (esq.) e 2, 5 e 6 (dir.)

Observa-se que a frequência das distâncias 2 e 3 aumenta com o passar do tempo e a evolução da rede, enquanto a ocorrência das demais distâncias diminui. O modelo indica aumento significativo da frequência da distância 3, e a extrapolação indica que mais de 50% das distâncias entre pares de vértices terão valor igual a 3 em julho de 2010. Observa-se também que a ocorrência de distâncias iguais a 6 tende a zero no período extrapolado.

O trabalho [Magoni and Pansiot 2001] indica que a distância média, o raio e o diâmetro da rede são invariantes, pois permaneceram estáveis no período analisado naquele trabalho. Nossos resultados indicam também a estabilidade das medidas de raio e diâmetro, mas indicam queda pequena porém consistente na distância média.

5.4. Evolução da Composição das Bordas da Rede

A evolução da rede em relação aos vértices nas bordas foi avaliada. Modelamos a evolução da ocorrência de nodos em árvores e o tamanho proporcional do subgrafo *2-core* em relação ao grafo completo da rede. A equação obtida para modelar a frequência de nodos em árvores é $0.262x^{-0.346}$ com conjunto de confiança igual a $(0, 262x^{-0.391}; 0, 262x^{-0.291})$. O modelo para a fração de nodos que pertence ao subgrafo *2-core* é dado pela equação de potência $0.745x^{0.077}$ com conjunto de confiança igual a $(0, 712x^{0.049}; 0, 778x^{0.106})$, com intervalo de confiança de 99%.

A Figura 11 apresenta modelos para a frequência relativa dos nodos em árvores nas bordas e a proporção do subgrafo *2-core*, bem como a extrapolação para o período de previsão. Observa-se que haverá cada vez menos subgrafos acíclicos nas bordas, o que evidencia o aumento da conectividade também nas bordas da rede.

6. Conclusão

Esse artigo apresentou uma análise da evolução da topologia da Internet no período de 2004 a 2007, e propôs modelos para as métricas que permitem a extrapolação e inferências sobre a evolução da topologia. Previsões de evolução da topologia no período de 2008 a 2010 foram apresentadas. Esse trabalho atualiza estudos anteriores sobre a topologia

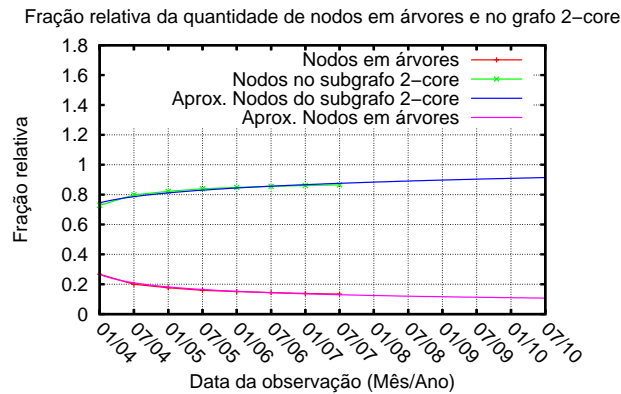


Figura 11. Evolução da freqüência de nodos em árvores e de nodos no subgrafo 2-core no período de janeiro de 2004 a 2010.

da Internet, encontrados na literatura para dados coletados até 2002, e mostra que alguns padrões de evolução se mantiveram e outros foram alterados. Os padrões de evolução que continuam sendo observados são a diminuição da distância média entre pares de vértices e o crescimento mais acelerado do número de arestas em relação ao número de vértices, com consequente aumento do grau médio. Os padrões alterados referem-se à taxa de crescimento da rede, que atualmente é menor do que as taxas observadas no passado, indicando que a evolução da rede iniciou uma nova etapa. Todos os padrões de evolução identificados evidenciam que a rede está cada vez mais compacta e conectada, tanto no centro quanto nas bordas. Os resultados e previsões podem ser úteis em várias áreas de estudo sobre a Internet, tais como protocolos de roteamento, simulações, geradores de topologias sintéticas, segurança e proteção contra ataques.

Agradecimentos

Este trabalho foi parcialmente financiado pelo CNPq. Agradecemos ao grupo do *Internet Research Laboratory* da UCLA pela coleta e por tornar disponível os dados da topologia da Internet na Web.

Referências

- Alderson, D., Li, L., Willinger, W., and Doyle, J. C. (2005). Understanding Internet Topology: principles, models, and validation. *IEEE/ACM Transactions on Networking*, 13(6):1205–1218.
- Bu, T. and Towsley, D. (2002). On Distinguishing between Internet Power-law Topology Generators. In *Proc. IEEE INFOCOM*, volume 2, pages 638– 647.
- Chang, H., Govindan, R., Jamin, S., Shenker, S. J., and Willinger, W. (2004). Towards Capturing Representative AS-level Internet Topologies. *Computer Networks*, 44(6).
- Chang, H. and Willinger, W. (2006). Difficulties Measuring the Internet’s AS-Level Ecosystem. In *Proceedings of IEEE INFOCOM*. Invited Paper.
- Chen, Q., Chang, H., Govindan, R., Jamin, S., Shenker, S., and Willinger, W. (2002). The Origin of Power Laws in Internet Topologies Revisited. In *Proc. IEEE INFOCOM*.

- Cohen, R. and Raz, D. (2006). The Internet Dark Matter – on the Missing Links in the AS Connectivity Map. In *Proceedings of IEEE INFOCOM*.
- Dorogovtsev, S. and Mendes, J. (2003). *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press.
- Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On Power-law Relationships of the Internet Topology. In *Proc. ACM SIGCOMM*, pages 251–262, New York, NY, USA. ACM Press.
- FIND (2007). Future Internet Design. <http://www.nets-find.net>.
- GENI (2007). Global Environment for Network Innovations. <http://www.geni.net>.
- He, Y., Siganos, G., Faloutsos, M., and Krishnamurthy, S. (2007). A Systematic Framework for Unearthing the Missing Links: Measurements and Impact. In *USENIX/SIGCOMM NSDI 2007*.
- IRL (2006). Internet Research Laboratory. <http://irl.cs.ucla.edu/topology>.
- Jaiswal, S., Rosenberg, A. L., and Towsley, D. (2004). Comparing the Structure of Power-law Graphs and the Internet AS Graph. In *Proc. 12th IEEE International Conference on Network Protocols (ICNP 2004)*.
- Leskovec, J., Kleinberg, J., and Faloutsos, C. (2007). Graph evolution: Densification and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1):2.
- Magoni, D. and Pansiot, J. J. (2001). Analysis of the Autonomous System Network Topology. *ACM SIGCOMM Comput. Commun. Rev.*, 31(3):26–37.
- Mahadevan, P., Krioukov, D., Fall, K., and A., V. (2006a). Systematic Topology Analysis and Generation Using Degree Correlations. *SIGCOMM 2006*.
- Mahadevan, P., Krioukov, D., Fomenkov, M., Dimitropoulos, X., k c claffy, and Vahdat, A. (2006b). The Internet AS-level Topology: three data sources and one definitive metric. *SIGCOMM Comput. Commun. Rev.*, 36(1):17–26.
- Oliveira, R. V., Zhang, B., and Zhang, L. (2007). Observing the Evolution of Internet AS Topology. *SIGCOMM Comput. Commun. Rev.*, 37(4):313–324.
- Palmer, C. R. and Steffan, J. G. (2000). Generating Network Topologies that Obey Power Laws. In *Proceedings of GLOBECOM 2000*.
- Park, S.-T., Pennock, D. M., and Giles, C. L. (2004). Comparing Static and Dynamic Measurements and Models of the Internet’s AS Topology. In *Proc. IEEE INFOCOM*.
- Pastor-Satorras, R. and Vespignani, A. (2004). *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press.
- RouteViews (2007). The Route Views Project. <http://www.antic.uoregon.edu/route-views/> (acessado em 2007).
- Tangmunarunkit, H., Govindan, R., Jamin, S., Shenker, S., and Willinger, W. (2002). Network Topology Generators: Degree-based vs. Structural. *ACM SIGCOMM Comput. Commun. Rev.*, 32(4):147–159.
- Zhang, B., Liu, R. A., Massey, D., and Zhang, L. (2005). Collecting the Internet AS-level Topology. *ACM Computer Communication Review*, 35(1):53–61.